2020

Adjustment criteria for recovering causal effects from missing data

Mojdeh Saadati
Iowa State University

Adjustment criteria for recovering causal effects from missing data

by

Mojdeh Saadati

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Computer Science

Program of Study Committee:
Jin Tian, Major Professor
Qi Li
Borzoo Bonakdarpour

Iowa State University

Ames, Iowa

2020

DEDICATION

I would like to dedicate this thesis to my family, my mother, father, sister, aunt and uncle. Definitely, without the help of any of them, I wasn't able to be where I am today.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Jin Tian for his guidance, patience and support throughout this research and the writing of this thesis. His insights and feedbacks have often inspired me and renewed my hopes for completing my graduate education. I would also like to thank my committee members for their contributions to this work: Dr. Qi Li and Dr. Borzoo Bonakdarpour. I would additionally like to thank Nicole Lewis, Samik Basu and Pavan Aduri for her guidance throughout my graduation career path and Dr. Hridish Rajan for making a welcoming environment in the department to pursue my degree.

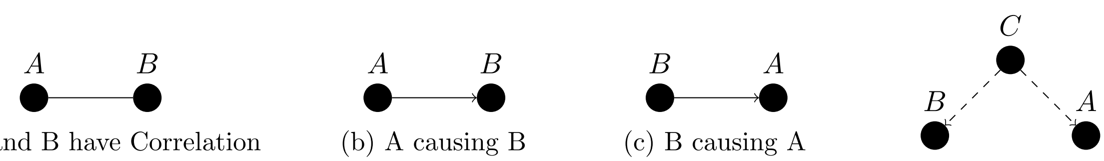
ABSTRACT

Confounding bias, missing data, and selection bias are three common obstacles to valid causal inference in the data sciences. Covariate adjustment is the most pervasive technique for recovering casual effects from confounding bias. In this thesis we introduce a covariate adjustment formulation for controlling confounding bias in the presence of missing-not-at-random data and develop a necessary and sufficient condition for recovering causal effects using the adjustment. We also introduce an adjustment formulation for controlling both confounding and selection biases in the presence of missing data and develop a necessary and sufficient condition for valid adjustment. Furthermore, we present an algorithm that lists all valid adjustment sets and an algorithm that finds a valid adjustment set containing the minimum number of variables,which are useful for researchers interested in selecting adjustment sets with desired properties.

CHAPTER 1. INTRODUCTION

Discovering causal relationships from observational data has been an important task in empirical sciences, for example, assessing the effect of a drug on curing diabetes, a fertilizer on growing agricultural products, and an advertisement on the success of a political party. To answer this causal questions, scientist usually conduct an experiment. In the experiment, a random sample is taken from the population, and then the sample is divided in two equal groups that have the same distribution for all variables except the experimental variables. Then the treatment is applied to one of the groups and favorable outcome variable is measured in both groups. The difference between outcome variable's value will refer to as the causal effects.

Conducting an experiment is not always possible due to ethical issues or limitations exist in collecting samples. The alternative approach to compute the causal effects is utilizing the data that already exist. The collected data can provide information about correlation. However, correlation does not imply causation. Consider the data set of two variables A and B. From data we can understand the Fig. 1.1(a). However, its not possible to distinguish between Fig. 1.1 {c, b, d}.



(a) A and B have Correlation (b) A causing B (c) B causing A (d) B causing A

Figure 1.1: An example for illustrating incapability of data in determining the causal effects

One major challenge to estimating the effect of a treatment on an outcome from observational data is the existence of *confounding bias*, which happens due to the existence of variables called confounders that affecting both treatments and outcomes, and triggers computed result be different than the causal relationship.

1.1 Identifiablity problem

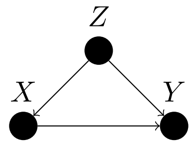
Recovering causal effects from observational data is formally addressed as the *identifiability problem* in (Pearl, 2009), which concerns with computing the effect of a set of treatment variables ($\mathbf{X}$) on a set of outcome variables ($\mathbf{Y}$), denoted by $P(\mathbf{y} \mid do(\mathbf{x}))$, given observed probability distribution $P(\mathbf{V})$ and a causal graph G, where $P(\mathbf{V})$ corresponds to the observational data and G is a directed acyclic graph (DAG) representing qualitative causal relationship assumptions between variables in the domain. This graph is mostly obtained based on background knowledge provided from variables. The effect $P(\mathbf{y} \mid do(\mathbf{x}))$ may not be equal to its probabilistic counterpart $P(\mathbf{y} \mid \mathbf{x})$ due to the existence of variables called *covariates* that are affecting both the treatments and outcomes and the difference is known as confounding bias.

For example, Fig. 1.2(a) shows a causal graph where variable Z is a covariate for estimating the effect of X on Y.

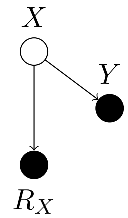
Confounding bias problem has been studied extensively in the field. In principle the identifiability problem can be solved using a set of causal inference rules called *do-calculus* (Pearl, 1995). Do-calculus is a set of three inference rules that can be applied to a causal effect expression $P(\mathbf{y} \mid do(\mathbf{x}))$ to obtain equivalent probability expressions of observed quantities. Complete identification algorithms have been developed-i.e., they are guaranteed to return equivalent set expressions based on observed distributions when the causal query is recoverable and return null when recoverability is not feasible (Tian and Pearl, 2002; Huang and Valtorta, 2006; Shpitser and Pearl, 2006). In practice, however, the most widely used method for controlling the confounding bias is the following "adjustment formula"

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{Z} = \mathbf{z}) P(\mathbf{Z} = \mathbf{z}),$$

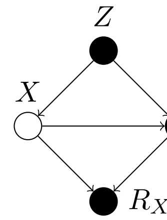
which dictates that the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ can be computed by *controlling* for a set of covariates $\mathbf{Z}$. Pearl provided a set of criterion called back-door under which a set $\mathbf{Z}$ makes the adjustment formula hold (Pearl, 1995).



(a) Confounding bias



(b) A m-graph encoding MNAR data

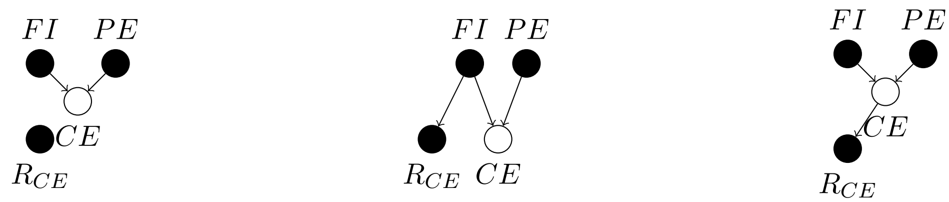


(c) Confounding bias in MNAR data

Figure 1.2: Examples for confounding bias in MNAR data

1.2 Missingness mechanism

Another major challenge to valid causal inference is the missing data problem, which occurs when some variable values are missing from observed data. Missing data is a common problem in empirical sciences. Indeed there is a large literature on dealing with missing data in diverse disciplines including statistics, economics, social sciences, and machine learning. To analyze data with missing values, it is imperative to understand the mechanisms that lead to missing data. The seminal work by Rubin (Rubin, 1976) classifies missing data mechanisms into three categories: *missing completely at random (MCAR)*, *missing at random (MAR)*, and *missing not at random (MNAR)*. MCAR and MAR are also called ingonable missing data and the data under MNAR is called non-ingorable missing data respectively. Roughly speaking, the mechanism is MCAR if whether variable values are missing is completely independent of the values of variables in the data set; the mechanism is MAR when missingness is independent of the missing values given the observed values; and the mechanism is MNAR if it is neither MCAR nor MAR. For example, assume that in a study of the effect of family income (FI) and parent's education level (PE) on the quality of child's education (CE), some respondents chose not to reveal their child's education quality for various reasons. Fig. 1.3 shows causal graphs representing the three missing data mechanisms where R_{CE} is an indicator variable such that $R_{CE} = 0$ if the CE value is missing and $R_{CE} = 1$ otherwise. In these graphs solid circles represent always-observed variables and hollow circles represent variables that could have missing values. The model in Fig. 1.3(a) is MCAR, e.g., respondents decide to reveal the child's education quality based on coin-flips. The model in Fig. 1.3(b) is MAR, where respondents with higher family income have a higher chance of revealing the child's education quality; however whether the CE values are missing is independent of the actual values of CE given the FI value. The model in Fig. 1.3(c) is MNAR, where respondents with higher child's education quality have a higher chance of revealing it, i.e., whether the CE values are missing depends on the actual values of CE.



(a) MCAR (b) MAR (c) MNAR

Figure 1.3: Three types of missing data mechanisms

It is known that when the data is MAR, the underlying distribution is estimable from observed data with missing values. Then a causal effect is estimable if it is identifiable from the observed distribution (Mohan and Pearl, 2014). However, if the data is MNAR, whether a probabilistic distribution or a causal effect is estimable from missing data depends closely on both the query and the exact missing data mechanisms. For example, in the MNAR model in Fig. 1.2(b), $P(X)$ cannot be estimated consistently even if infinite amount of data are collected, while $P(y|do(x)) = P(y|x) = P(y|x, R_X = 1)$ is estimable from missing data. On the other hand, in the MNAR model in Fig. 1.2(c), $P(y|do(x))$ is not estimable. In the MNAR model in Fig. 1.3(c), neither $P(CE)$ nor $P(CE \mid do(FI))$ can be estimated from observed data with missing values.

1.3 Identifiability under missing data

Various techniques have been developed to deal with missing data in statistical inference which leads to well-known statistical software such as R, SPSS, STATA and SAS dedicate packages for handling missing data. Listwise deletion (Little and Rubin, 1986) is one of the most common techniques in handling missing data that can generate unbiased estimate under MCAR assumption. There are some cases that this technique is able to produce close estimate in MAR and MNAR assumption. Multiple imputation (Rubin, 1978) and Maximum Likelihood estimation (Koller et al., 2009) are two approaches for handling missing data in MAR assumption. When the among of data that is going to be removed by Listwise deletion is significant, these two methods can be utilized which are producing approximately consistent and unbiased estimate for MAR data (Allison, 2001). Multivariate imputation by chained MissForest is an iterative algorithm based on randomForest algorithm (Breiman, 2001). In each iteration, this algorithm assigns randomForests for each variables' observed values and predict a missing entries based on it. It stops when a certain criterion is statisfied. (Stekhoven and Bühlmann, 2011) Several techniques have been developed for handling MNAR such as Pattern Mixture models and Heckman's model, also named sample selection model Galimard et al (2018) and Galimard et al (2016), or ampute, this technique assume independency between X and R_X, however these techniques consider specific assumptions for MNAR structure that is not always true and lacks generalizing to all types of MNAR structures.

The use of graphical models called *m-graphs* for inference with missing data was more recent (Mohan et al., 2013). M-graphs provide a general framework for inference with arbitrary types of missing data mechanisms including MNAR. Sufficient conditions for determining whether probabilistic queries (e.g., $P(\mathbf{y} \mid \mathbf{x})$ or $P(\mathbf{x}, \mathbf{y})$) are estimable from missing data are provided in (Mohan et al., 2013; Mohan and Pearl, 2014). General algorithms for identifying the joint distribution have been developed in (Shpitser et al., 2015; Tian, 2017).

The problem of identifying causal effects $P(\mathbf{y} \mid do(\mathbf{x}))$ from missing data in the causal graphical model settings has not been well studied. To the best of our knowledge the only results are the sufficient conditions given in (Mohan and Pearl, 2014). The goal of this paper is to provide general conditions under which the causal effects can be identified from missing data using the covariate adjustment formula, which is the most pervasive method in practice for causal effect estimation under confounding bias.

1.4 Identifiablity under missing data and selection bias

Identifying causal effects from selection bias has been studied in the literature (Bareinboim et al., 2014; Bareinboim and Tian, 2015). Adjustment formulas for recovering causal effects under selection bias have been introduced and complete graphical criteria have been developed (Correa and Bareinboim, 2017; Correa et al., 2018). However these results are not applicable to the missing data problems which have much richer missingness patterns than could be modeled by selection bias. To the best of our knowledge, using adjustment for causal effect identification when the observed data suffers from missing values or both selection bias and missing values has not been studied in the causal graphical model settings. In this thesis we will provide two adjustment formula, m-adjustment and ms-adjustment, that by taking advantages of sets of variables in the data set compute the causal effects in missing data and mention the criteria that those set of variables should satisfy to be usable in the data. We then will mention some potential obstacles that exist in utilizing these formula and propose an alternative approach to handle those obstacles. At last, two algorithms developed to return all sets of variables that satisfy the criteria and the minimum size set solution.

CHAPTER 2. DEFINITIONS AND RELATED WORK

2.1 Definition

Each variable will be represented with a capital letter (X) and its realized value with the small letter (x). We will use bold letters ($\mathbf{X}$) to denote sets of variables.

2.1.1 D-separation

Given sets of variables $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ in a causal graph G we say $\mathbf{X}$ is d-separated from $\mathbf{Y}$ by $\mathbf{Z}$ if all paths from $\mathbf{X}$ to $\mathbf{Y}$ are blocked by variables in $\mathbf{Z}$ based on the d-separation rule-i.e., a path p is blocked or d-separated by $\mathbf{Z}$ if there is a chain ($i \rightarrow m \rightarrow j$) or fork ($i \leftarrow m \rightarrow j$) in path p that $m \in \mathbf{Z}$ or collider ($i \rightarrow m \leftarrow j$) that $m \notin \mathbf{Z}$.

It is proved that two variables are independent in a probability distribution if and only if they are d-separated in the causal graph corresponding to that distribution, Pearl (1995).

2.1.2 Causal models

The systematic analysis of confounding bias and missing data mechanisms requires a formal language where the characterization of the underlying data-generating model can be encoded explicitly. We use the language of Structural Causal Models (SCM) (Pearl, 2009).

In SCMs, performing an action/intervention of setting $\mathbf{X}{=}\mathbf{x}$ is represented through the do-operator, $do(\mathbf{X}{=}\mathbf{x})$, which induces an experimental distribution $P(\mathbf{y}|do(\mathbf{x}))$, known as the causal effect of $\mathbf{X}$ on $\mathbf{Y}$. For a detailed discussion of SCMs, we refer readers to (Pearl, 2009).

Each SCM M has a causal graph G associated to it, with directed arrows encoding direct causal relationships and dashed-bidirected arrows encoding the existence of an unobserved common causes (e.g., see Fig. 3.1 in the next chapter). We use typical graph-theoretic terminology

$Pa(\mathbf{C}), Ch(\mathbf{C}), De(\mathbf{C}), An(\mathbf{C})$ representing the union of $\mathbf{C}$ and respectively the parents, children, descendants, and ancestors of $\mathbf{C}$. We use $G_{\overline{\mathbf{C_1}}\underline{\mathbf{C_2}}}$ to denote the graph resulting from deleting all incoming edges to $\mathbf{C}_1$ and all outgoing edges from $\mathbf{C}_2$ in G. The expression $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G$ denotes that $\mathbf{X}$ is d-separated from $\mathbf{Y}$ given $\mathbf{Z}$ in the corresponding causal graph G (Pearl, 2009)(subscript G may be omitted).

Do-calculus is introduced for the first time by Pearl (1995) is a set of three rules to derive causal expressions from other causal quantities, so we can have all equivalent causal quantities. The rules are as follows:

Rule 1(Insertion/deletion of observation):

$$P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{w}) \ if \ (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}}} \tag{2.1}$$

Rule 2(Action/observation exchange):

$$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{w}) \ if \ (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X},\underline{Z}}} \tag{2.2}$$

Rule 3(Insertion/deletion action):

$$P(\mathbf{y} \mid do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = \sum_{\mathbf{z}} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{w}) \ if \ (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X},\overline{Z(W)}}} \tag{2.3}$$

Where Z(W) is a set of $\mathbf{Z}$ nodes that are not ancestor of any node in $\mathbf{W}$ in the graph $G_{\overline{X}}$.

2.1.3 Missing data and m-graphs

To deal with missing data, we use *m-graphs* introduced in (Mohan et al., 2013) to represent both the data generation model and the missing data mechanisms. M-graphs enhance the causal graph G by introducing a set $\mathbf{R}$ of binary missingness indicator variables. We will also partition the set of observable variables $\mathbf{V}$ into $\mathbf{V}_o$ and $\mathbf{V}_m$ such that $\mathbf{V}_o$ is the set of variables that will be observed in all data cases and $\mathbf{V}_m$ is the set of variables that are missing in some data cases and observed in other cases. Every variable $V_i \in \mathbf{V}_m$ is associated with a variable $R_{V_i} \in \mathbf{R}$ such that, in any observed data case, $R_{V_i} = 0$ if the value of corresponding V_i is missing and $R_{V_i} = 1$ if V_i is observed. We assume that $\mathbf{R}$ variables may not be parents of variables in $\mathbf{V}$, since $\mathbf{R}$ variables are missingness indicator variables and we assume that the data generation process over $\mathbf{V}$ variables does not depend on the missingness mechanisms. For any set $\mathbf{C} \subseteq \mathbf{V}_m$, let $\mathbf{R_C}$ represent the set of $\mathbf{R}$ variables corresponding to variables in $\mathbf{C}$. See Fig. 1.3 for examples of m-graphs, in which we use solid circles to represent always observed variables in $\mathbf{V}_o$ and $\mathbf{R}$, and hollow circles to represent partially observed variables in $\mathbf{V}_m$.

2.2 Related work

2.2.1 Causal effect identification by adjustment

Covariate adjustment is the most widely used technique for identifying causal effects from observational data. Formally, it introduces as Adjustment Formula (Pearl, 2009).

Definition 1 *Given a causal graph G over a set of variables $\boldsymbol{V}$, a set $\boldsymbol{Z}$ is called* covariate adjustment *(or adjustment for short) for estimating the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, if, for any distribution $P(\boldsymbol{V})$ compatible with G, it holds that*

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}) P(\boldsymbol{z}). \tag{2.4}$$

Pearl developed the celebrated "Backdoor Criterion" to determine whether a set is admissible for adjustment (Pearl, 1995) given in the following:

Definition 2 *A set of variables $\boldsymbol{Z}$ satisfies the backdoor criterion relative to a pair of variables $(\boldsymbol{X}, \boldsymbol{Y})$ in a causal graph G if:*

a) No node in $\boldsymbol{Z}$ is a descendant of $\boldsymbol{X}$, and

b) $\boldsymbol{Z}$ blocks every path between $\boldsymbol{X}$ and $\boldsymbol{Y}$ that contains an arrow into $\boldsymbol{X}$.

Complete graphical conditions have been derived for determining whether a set is admissible for adjustment (Shpitser et al., 2010; van der Zander et al., 2014; Perkovic et al., 2017) as follows.

Definition 3 (Proper Causal Path) *A proper causal path from a node* $X \in \boldsymbol{X}$ *to a node* $Y \in \boldsymbol{Y}$ *is a causal path (i.e., a directed path) which does not intersect* $\boldsymbol{X}$ *except at the beginning of the path.*

Definition 4 (Adjustment Criterion, Shpitser et al. (2010)) *A set of variables* $\boldsymbol{Z}$ *satisfies the adjustment criterion relative to a pair of variables* $(\boldsymbol{X}, \boldsymbol{Y})$ *in a causal graph* G *if:*

a) No element of $\boldsymbol{Z}$ *is a descendant in* $G_{\overline{X}}$ *of any* $W \notin \boldsymbol{X}$ *which lies on a proper causal path from* $\boldsymbol{X}$ *to* $\boldsymbol{Y}$*.*

b) All non-causal paths between $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *in* G *are blocked by* $\boldsymbol{Z}$*.*

A set $\mathbf{Z}$ is an admissible adjustment for estimating the causal effect of $\mathbf{X}$ on $\mathbf{Y}$ by the adjustment formula if and only if it satisfies the adjustment criterion.

CHAPTER 3. ADJUSTMENT FOR RECOVERING CAUSAL EFFECTS FROM MISSING DATA

In this section we address the task of recovering a causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ from missing data given a m-graph G over observed variables $\mathbf{V} = \mathbf{V}_o \cup \mathbf{V}_m$ and missingness indicators $\mathbf{R}$. The main difference with the well studied identifiability problem (Pearl, 2009), where we attempt to identify $P(\mathbf{y} \mid do(\mathbf{x}))$ from the joint distribution $P(\mathbf{V})$, lies in that, given data corrupted by missing values, $P(\mathbf{V})$ itself may not be recoverable. Instead, a distribution like $P(\mathbf{V}_o, \mathbf{V}_m, \mathbf{R} = 1)$ is assumed to be estimable from observed data cases in which all variables in $\mathbf{V}$ are observed (i.e., complete data cases). In general, in the context of missing data, the probability distributions in the form of $P(\mathbf{V}_o, \mathbf{W}, \mathbf{R}_\mathbf{W} = 1)$ for any $\mathbf{W} \subseteq \mathbf{V}_m$, called *manifest distributions*, are assumed to be estimable from observed data cases in which all variables in $\mathbf{W}$ are observed (values of variables in $\mathbf{V}_m \setminus \mathbf{W}$ are possibly missing). The problem of recovering probabilistic queries from the manifest distributions has been studied in (Mohan et al., 2013; Mohan and Pearl, 2014; Shpitser et al., 2015; Tian, 2017).

3.1 Adjustment under missing data

In this section, we will extend the adjustment formula for identifying causal effects to the context of missing data.

3.1.1 M-adjustment formula

We will extend the adjustment formula for identifying causal effects to the context of missing data based on the following observation which is stated in Theorem 1 in (Mohan et al., 2013):

Lemma 1 *For any* $\boldsymbol{W}_o, \boldsymbol{Z}_o \in \boldsymbol{V}_o$ *and* $\boldsymbol{W}_m, \boldsymbol{Z}_m \in \boldsymbol{V}_m$, $P(\boldsymbol{W}_o, \boldsymbol{W}_m \mid \boldsymbol{Z}_o, \boldsymbol{Z}_m, \boldsymbol{R}_{\boldsymbol{W}_m \cup \boldsymbol{Z}_m} = 1)$ *is recoverable.*

Formally, we introduce the adjustment formula for recovering causal effects from missing data by extending Eq. (2.4) as follows.

Definition 5 (M-Adjustment Formula) *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$, a set $\boldsymbol{Z} \subseteq \boldsymbol{V}$ is called a m-adjustment (adjustment under missing data) set for estimating the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$, if, for every model compatible with G, it holds that*

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{R}_{\boldsymbol{W}} = 1) P(\boldsymbol{z} \mid \boldsymbol{R}_{\boldsymbol{W}} = 1), \tag{3.1}$$

where $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$.

In the above formulation, we allow that the treatments $\mathbf{X}$, outcomes $\mathbf{Y}$, and covariates $\mathbf{Z}$ all could contain $\mathbf{V}_m$ variables that have missing values. Both terms on the right-hand-side of Eq. (3.1) are recoverable based on Lemma 1. Therefore the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ is recoverable if it can be expressed in the form of m-adjustment.

We look for conditions under which a set $\mathbf{Z}$ is admissible as m-adjustment. In principle this can be derived using do-calculus. As an example, consider the m-graph in Fig. 3.1 where R_1, R_2, R_3 are missingness indicators for Z_{m1}, Z_{m2}, Z_{m3} respectively. We show that $\{Z_{m1}\}$ is m-adjustment admissible for recovering $P(y \mid do(x_1, x_2))$ using do-calculus derivation as follows:

$$P(y \mid do(x_1, x_2))$$

$$= P(y \mid do(x_1, x_2), R_1 = 1) \tag{3.2}$$

$$= \sum_{Z_{m1}} P(y \mid do(x_1, x_2), Z_{m1}, R_1 = 1) P(Z_{m1} \mid R_1 = 1, do(x_1, x_2)) \tag{3.3}$$

$$= \sum_{Z_{m1}} P(y \mid do(x_1, x_2), Z_{m1}, R_1 = 1) P(Z_{m1} \mid R_1 = 1) \tag{3.4}$$

$$= \sum_{Z_{m1}} P(y \mid x_1, x_2, Z_{m1}, R_1 = 1) P(Z_{m1} \mid R_1 = 1). \tag{3.5}$$

In general using do-calculus to recover causal effects is difficult due to many possible ways of applying do-calculus rules in every stage of the derivation. Intuitively, we can start with the adjustment formula (2.4), consider an adjustment set as a candidate m-adjustment set, and

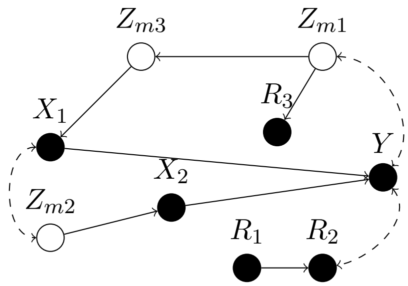


Figure 3.1: An example of m-adjustment in a MNAR model

then check for needed conditional independence relations. Based on this intuition, we obtain a straightforward sufficient condition for a set $\mathbf{Z}$ to be a m-adjustment set as follows.

Proposition 1 *A set $\boldsymbol{Z}$ is a m-adjustment set for estimating the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ if, letting $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$,*

a) $\boldsymbol{Z}$ satisfies the adjustment criterion (Def. 4),

b) $\boldsymbol{R_W}$ is d-separated from $\boldsymbol{Y}$ given $\boldsymbol{X}$, $\boldsymbol{Z}$, i.e., $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R_W} \mid \boldsymbol{X}, \boldsymbol{Z})$,

c) $\boldsymbol{Z}$ is d-separated from $\boldsymbol{R_W}$, i.e., $(\boldsymbol{Z} \perp\!\!\!\perp \boldsymbol{R_W})$.

Condition (a) makes sure that the causal effect can be identified in terms of the adjustment formula (2.4). Then given Conditions (b) and (c), Eq. (2.4) is equal to Eq. (3.1). For example, in Fig. 3.1, $\{Z_{m1}\}$, $\{Z_{m3}\}$, and $\{Z_{m1}, Z_{m3}\}$ all satisfy the back-door criterion (and therefore the adjustment criterion), however only $\{Z_{m1}\}$ satisfies the conditions in Proposition 1 ($\{Z_{m3}\}$, and $\{Z_{m1}, Z_{m3}\}$ do not satisfy Condition (c) because Z_{m3} is not d-separated from R_3).

However this straightforward criterion in Proposition 1 is not necessary. To witness, consider the set $\{V_{m1}, V_{m2}\}$ in Fig. 3.2 which satisfies the back-door criterion but not the conditions in Proposition 1 because V_{m2} is not d-separated from R_2. Still, it can be shown that $\{V_{m1}, V_{m2}\}$ is a

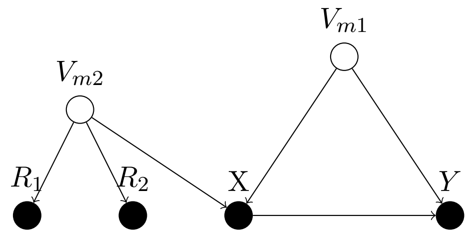


Figure 3.2: In this m-graph V_{m2} is not d-separated from R_2. However, $\{V_{m2}, V_{m1}\}$ is an admissible m-adjustment set.

m-adjustment set by do-calculus derivation as follows:

$$P(y \mid do(x))$$

$$= P(y \mid do(x), R_1 = 1, R_2 = 1) \tag{3.6}$$

$$= \sum_{v_{m1}, v_{m2}} P(y \mid do(x), v_{m1}, v_{m2}, R_1 = 1, R_2 = 1) P(v_{m1}, v_{m2} \mid R_1 = 1, R_2 = 1, do(x)) \tag{3.7}$$

$$= \sum_{v_{m1}, v_{m2}} P(y \mid do(x), v_{m1}, v_{m2}, R_1 = 1, R_2 = 1) P(v_{m1}, v_{m2} \mid R_1 = 1, R_2 = 1) \tag{3.8}$$

$$= \sum_{v_{m1}, v_{m2}} P(y \mid x, v_{m1}, v_{m2}, R_1 = 1, R_2 = 1) P(v_{m1}, v_{m2} \mid R_1 = 1, R_2 = 1). \tag{3.9}$$

3.1.2 M-adjustment criterion

Next we introduce a complete criterion to determine whether a covariate set is admissible as m-adjustment to recover causal effects from missing data, extending the existing work on adjustment (Shpitser et al., 2010; van der Zander et al., 2014; Correa and Bareinboim, 2017; Correa et al., 2018; Perkovic et al., 2017).

Definition 6 (M-Adjustment Criterion) *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$, and disjoint sets of variables $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z} \subseteq \boldsymbol{V}$, letting $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$, $\boldsymbol{Z}$ satisfies the m-adjustment criterion relative to the pair $(\boldsymbol{X}, \boldsymbol{Y})$ if*

a) No element of $\boldsymbol{Z}$ is a descendant in $G_{\overline{X}}$ of any $W \notin \boldsymbol{X}$ which lies on a proper causal path from $\boldsymbol{X}$ to $\boldsymbol{Y}$.

b) All non-causal paths between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in G are blocked by $\boldsymbol{Z}$ and $\boldsymbol{R}_{\boldsymbol{W}}$.

c) $\boldsymbol{R_W}$ *is d-separated from* $\boldsymbol{Y}$ *given* $\boldsymbol{X}$ *under the intervention of* $do(\boldsymbol{x})$*, i.e.,* $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R_W} \mid \boldsymbol{X})_{G_{\overline{X}}}$.

d) Every $X \in \boldsymbol{X}$ *is either a non-ancestor of* $\boldsymbol{R_W}$ *or it is d-separated from* $\boldsymbol{Y}$ *in* $G_{\underline{X}}$*, i.e.,* $\forall X \in \boldsymbol{X} \cap An(\boldsymbol{R_W}), (X \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$.

Theorem 1 (M-Adjustment) *A set* $\boldsymbol{Z}$ *is a m-adjustment set for recovering causal effect of* $\boldsymbol{X}$ *on* $\boldsymbol{Y}$ *by the m-adjustment formula in Def. 5 if and only if it satisfies the m-adjustment criterion in Def. 6.*

The proof of Theorem 1 is presented in the Appendix B. Conditions (a) and (b) in Def. 6 echo the adjustment criterion in Def. 4 and it can be shown that if $\mathbf{Z}$ satisfies the m-adjustment criterion then it satisfies the adjustment criterion (using the fact that no variables in $\mathbf{R}$ can be parents of variables in $\mathbf{V}$). In other words, we only need to look for m-adjustment sets from admissible adjustment sets.

As an example consider Fig. 3.2. Both $\{V_{m1}\}$ and $\{V_{m1}, V_{m2}\}$ satisfy the m-adjustment criterion (and the adjustment criterion too). According to Theorem 1, $P(y \mid do(x))$ can be recovered from missing data by m-adjustment given in Eq. (3.9), and can also by recovered as

$$P(y \mid do(x)) = \sum_{v_{m1}} p(y \mid x, v_{m1}, R_1 = 1) P(v_{m1} \mid R_1 = 1). \tag{3.10}$$

The algorithm for listing all valid m-adjustment sets and finding the minimum size m-adjustment set has been developed in (**?**)

3.2 Adjustment from both selection bias and missing data

In Sections 3.1 and 4 we have addressed the task of recovering causal effects by adjustment from missing data. In practice another common issue that data scientists face in estimating causal effects is selection bias. Selection bias can be modeled by introducing a binary indicator variable S such that $S = 1$ if a unit is included in the sample, and $S = 0$ otherwise Bareinboim et al. (2014). Graphically selection bias is modeled by a special hollow node S (drawn round with double border) that is pointed to by every variable in $\mathbf{V}$ that affects the process by which an unit is included in the data. In Fig. 3.3(a), for example, selection is affected by the treatment variable X.

(a) Selection bias

(b) MNAR model with selection bias

ID	X	R_X	Y	S
1	1	1	0	1
2	0	1	1	1
3	NA	NA	NA	0
4	NA	0	1	1
5	NA	NA	NA	0

(c) An example data set compatible with the model in Fig. 3.3(b)

Figure 3.3: Examples of selection bias and MNAR

In the context of selection bias, the observed distribution is $P(\mathbf{V} \mid S = 1)$, collected under seletion bias, instead of $P(\mathbf{V})$. The goal of inference is to recover the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ from $P(\mathbf{V} \mid S = 1)$. The use of adjustment for recovering causal effects in this setting has been studied and complete adjustment conditions have been developed in Correa and Bareinboim (2017); Correa et al. (2018).

What if the observed data suffers from both selection bias and missing values? In the model in Fig. 3.3(b), for example, whether a unit is included in the sample depends on the value of the outcome Y. If a unit is included in the sample, the values of treatment X could be missing depending on the actual X values. Fig. 3.3(c) shows an example data set compatible with the model in Fig. 3.3(b) illustrating the difference between selection bias and missing data. To the best of our knowledge, causal inference under this setting has not been formally studied.

3.2.1 MS-adjustment formula

In this section, we will characterize the use of adjustment for causal effect identification when the observed data suffers from both selection bias and missing values. First we introduce an adjustment formula called *MS-adjustment* for recovering causal effect under both missing data and selection bias. Then we provide a complete condition under which a set $\mathbf{Z}$ is valid as MS-adjustment set.

Definition 7 *Given a m-graph G over observed variables $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ and missingness indicators $\boldsymbol{R}$ augmented with a selection bias indicator S, a set $\boldsymbol{Z} \subseteq \boldsymbol{V}$ is called a ms-adjustment (adjustment* *under missing data and selection bias) set for estimating the causal effect of* $\boldsymbol{X}$ *on* $\boldsymbol{Y}$*, if for every model compatible with* G *it holds that*

$$P(\boldsymbol{y} \mid do(\boldsymbol{x})) = \sum_{\boldsymbol{z}} P(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{R_W} = 1, S = 1) P(\boldsymbol{z} \mid \boldsymbol{R_W} = 1, S = 1), \tag{3.11}$$

where $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$.

Both terms on the right-hand-side of Eq. (3.11) are recoverable from selection biased data in which all variables in $\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$ are observed. Therefore the causal effect $P(\mathbf{y} \mid do(\mathbf{x}))$ is recoverable if it can be expressed in the form of ms-adjustment.

3.2.2 MS-adjustment criterion

Next we provide a complete criterion to determine whether a set $\mathbf{Z}$ is an admissible ms-adjustment.

Definition 8 *Given a m-graph* G *over observed variables* $\boldsymbol{V} = \boldsymbol{V}_o \cup \boldsymbol{V}_m$ *and missingness indicators* $\boldsymbol{R}$ *augmented with a selection bias indicator* S*, and disjoint sets of variables* $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$*, letting* $\boldsymbol{W} = \boldsymbol{V}_m \cap (\boldsymbol{X} \cup \boldsymbol{Y} \cup \boldsymbol{Z})$*,* $\boldsymbol{Z}$ *satisfies the ms-adjustment criterion relative to the pair* $(\boldsymbol{X}, \boldsymbol{Y})$ *if*

a) No element of $\boldsymbol{Z}$ *is a descendant in* $G_{\overline{X}}$ *of any* $W \notin \boldsymbol{X}$ *which lies on a proper causal path from* $\boldsymbol{X}$ *to* $\boldsymbol{Y}$*.*

b) All non-causal paths between $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *in* G *are blocked by* $\boldsymbol{Z}$*,* $\boldsymbol{R_W}$*, and* S*.*

c) $\boldsymbol{R_W}$ *and* S *are d-separated from* $\boldsymbol{Y}$ *given* $\boldsymbol{X}$ *under the intervention of* $do(\boldsymbol{x})$*. i.e.,* $(\boldsymbol{Y} \perp\!\!\!\perp (\boldsymbol{R_W} \cup S) \mid \boldsymbol{X})_{G_{\overline{X}}}$

d) Every $X \in \boldsymbol{X}$ *is either a non-ancestor of* $\{\boldsymbol{R_W}, S\}$ *or it is d-separated from* $\boldsymbol{Y}$ *in* $G_{\underline{X}}$*. i.e.,* $\forall X \in \boldsymbol{X} \cap An(\boldsymbol{R_W} \cup S), (X \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$*.*

Theorem 2 (MS-Adjustment) *A set* $\boldsymbol{Z}$ *is a ms-adjustment set for recovering causal effect of* $\boldsymbol{X}$ *on* $\boldsymbol{Y}$ *by the ms-adjustment formula in Definition 7 if and only if it satisfies the ms-adjustment criterion in Definition 8.*

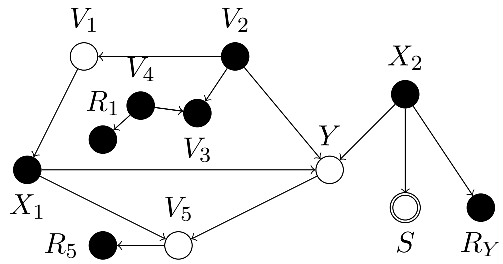


Figure 3.4: An example for recovering causal effect under both selection bias and MNAR data

To demonstrate the application of Theorem 2, consider the causal graph in Fig. 3.4 where V_1 ,V_5, Y may have missing values and the selection S depends on the values of X_2. To recover the causal effect of $\{X_1, X_2\}$ on variable Y, V_1 satisfies the ms-adjustment criterion. We obtain $P(y \mid do(x_1, x_2)) = \sum_{V_1} P(y \mid x_1, x_2, V_1, S = 1, R_y = 1, R_1 = 1)P(V_1 \mid S = 1, R_y = 1, R_1 = 1)$.

We note that the two algorithms given in Section 4, for listing all m-adjustment sets and finding a minimum size m-adjustment set, can be extended to list all ms-adjustment sets and find a minimum ms-adjustment set with minor modifications.

CHAPTER 4. LISTING M-ADJUSTMENT SETS

In the previous section we provided a criterion under which a set of variables $\mathbf{Z}$ is an admissible m-adjustment set for recovering a causal effect. It is natural to ask how to find an admissible set. In reality, it is common that more than one set of variables are admissible. In such situations it is possible that some m-adjustment sets might be preferable over others based on various aspects such as feasibility, difficulty, and cost of collecting variables. Next we first present an algorithm that systematically lists all m-adjustment sets and then present an algorithm that finds a minimum m-adjustment set. These algorithms provide flexibility for researchers to choose their preferred adjustment set based on their needs and assumptions.

4.1 Listing all admissible sets

It turns out in general there may exist exponential number of m-adjustment sets. To illustrate, we look for possible m-adjustment sets in the m-graph in Fig. 4.1 for recovering the causal effect $p(\mathbf{y} \mid do(\mathbf{x}))$ (this graph is adapted from a graph in Correa et al. (2018)). A valid m-adjustment set $\mathbf{Z}$ needs to close all the k non-causal paths from X to Y. $\mathbf{Z}$ must contain at least one variable in $\{V_{i1}, V_{i2}, V_{i3}\}$ for each $i = 1, \ldots, k$. Therefore, to close each path, there are 7 possible $\mathbf{Z}$ sets, and for k paths, we have total 7^k $\mathbf{Z}$ sets as potential m-adjustment sets. For each of them, Conditions (c) and (d) in Def. 6 are satisfied because $(\mathbf{R} \perp\!\!\!\perp Y \mid X)_{G_{\overline{X}}}$ and X is not an ancestor of any $\mathbf{R}$ variables. We obtain that there are at least 7^k number of m-adjustment sets.

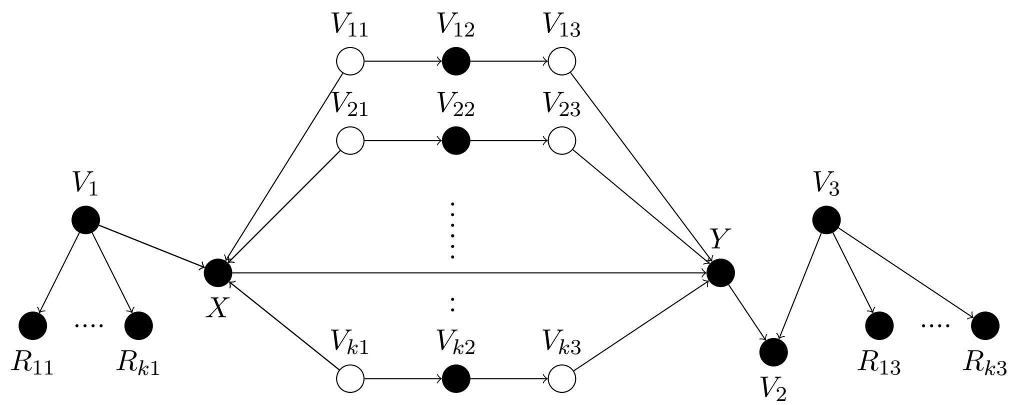


Figure 4.1: An example of exponential number of m-adjustment sets

The above example demonstrates that any algorithm that lists all m-adjustment sets can be computed in an exponential time. To deal with this issue, we will provide an algorithm with polynomial delay complexity Takata (2010). Polynomial delay algorithms require polynomial time to generate the first output (or indicate failure) and the time between any two consecutive outputs is polynomial as well.

To facilitate the construction of a listing algorithm, we introduce a graph transformation called *Proper Backdoor Graph* originally introduced by Van der Zander, Liskiewicz, and Textor (2014).[Proper Backdoor Graph van der Zander et al. (2014)] Let G be a causal graph, and $\mathbf{X}, \mathbf{Y}$ be disjoint subsets of variables. The proper backdoor graph, denoted as $G^{pbd}_{\mathbf{X},\mathbf{Y}}$, is obtained from G by removing the first edge of every proper causal path from $\mathbf{X}$ to $\mathbf{Y}$.

Next we present an alternative equivalent formulation of the m-adjustment criterion in Def. 6 that will be useful in constructing a listing algorithm. [M-Adjustment Criterion, Math. Version] Given a m-graph G over observed variables $\mathbf{V} = \mathbf{V}_o \cup \mathbf{V}_m$ and missingness indicators $\mathbf{R}$, and disjoint sets of variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$, letting $\mathbf{W} = \mathbf{V}_m \cap (\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z})$, $\mathbf{Z}$ satisfies the m-adjustment criterion relative to the pair $(\mathbf{X}, \mathbf{Y})$ if

a) $\mathbf{Z} \cap Dpcp(\mathbf{X}, \mathbf{Y}) = \phi$

b) $(\mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}, \mathbf{R_W})_{G^{pbd}_{\mathbf{X},\mathbf{Y}}}$

c) $(\mathbf{Y} \perp\!\!\!\perp \mathbf{R_W} \mid \mathbf{X})_{G_{\overline{X}}}$

d) $((\mathbf{X} \cap An(\mathbf{R_W})) \perp\!\!\!\perp \mathbf{Y})_{G_{\underline{X}}}$

where

$$D_{pcp}(\mathbf{X}, \mathbf{Y}) = De((De(\mathbf{X})_{G_{\overline{X}}} \setminus \mathbf{X}) \cap An(\mathbf{Y})_{G_{\overline{X}}}). \tag{4.1}$$

In Definition 4.1, $D_{pcp}(\mathbf{X}, \mathbf{Y})$, originally introduced in van der Zander et al. (2014), represents the set of descendants of those variables in a proper causal path from $\mathbf{X}$ to $\mathbf{Y}$.

Proposition 2 *Definition 4.1 and Definition 6 are equivalent.*

Note that all the proofs of the propositions and theorems in Section 4 are given in Appendix A.

Finally to help understanding the logic of the algorithm we introduce a definition originally introduced in Correa et al. (2018): [Family of Separators Correa et al. (2018)] For a disjoint set of variables $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{E}$ and $\mathbf{I} \subseteq \mathbf{E}$, a family of separators is defined as follows:

$$Z_{G(\mathbf{X},\mathbf{Y})}\langle \mathbf{I}, \mathbf{E} \rangle := \{\mathbf{Z} \mid (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G \text{ and } \mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{E}\}, \tag{4.2}$$

which represent the set of all sets that d-separate $\mathbf{X}$ and $\mathbf{Y}$ and encompass all variables in set $\mathbf{I}$ but do not have any variables outside $\mathbf{E}$.

Algorithm 1 presents the function ListMAdj that lists all the m-adjustment sets in a given m-graph G for recovering the causal effect of $\mathbf{X}$ on $\mathbf{Y}$. We note that the algorithm uses an external function FindSep described in van der Zander et al. (2014) (not presented in this thesis). FindSep(G, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{I}$, $\mathbf{E}$) will return a set in $Z_{G(\mathbf{X},\mathbf{Y})}\langle \mathbf{I}, \mathbf{E} \rangle$ if such a set exists; otherwise it returns $\perp$ representing failure.

Function ListMAdj works by first excluding all variables lying in the proper causal paths from consideration (Line 3) and then calling the function ListSepConditions (Line 4) to return all the m-adjustment sets. The function of ListSepConditions is summarized in the following proposition:

Proposition 3 (Correctness of ListSepConditions) *Given a m-graph G and sets of disjoint variables $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{E}$ and $\boldsymbol{I} \subseteq \boldsymbol{E}$, ListSepConditions lists all $\boldsymbol{Z}$ variables such that:*

Algorithm 1: Listing all the m-adjustment sets

1 **Function** *ListMAdj* $(G, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{V}_o, \boldsymbol{V}_m, \boldsymbol{R})$
2 $\quad G^{pbd}_{\mathbf{X},\mathbf{Y}} \leftarrow$ compute proper back-door graph G
3 $\quad \mathbf{E} \leftarrow (\mathbf{V}_o \cup \mathbf{V}_m \cup \mathbf{R}) \setminus \{\mathbf{X} \cup \mathbf{Y} \cup D_{pcp}(\mathbf{X},\mathbf{Y})\}$.
4 $\quad$ ListSepConditions$(G^{pbd}_{\mathbf{X},\mathbf{Y}}, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I} = \{\mathbf{R}_{\mathbf{X}\cap\mathbf{V}_m} \cup \mathbf{R}_{\mathbf{Y}\cap\mathbf{V}_m}\}, \mathbf{E})$
5 **Function** *ListSepConditions* $(G, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{R}, \boldsymbol{V}_o, \boldsymbol{V}_m, \boldsymbol{I}, \boldsymbol{E})$
6 $\quad$ **if** $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R_I} \mid \boldsymbol{X})_{G_{\overline{\boldsymbol{X}}}}$ *and* $((\boldsymbol{X} \cap An(\boldsymbol{R_I})) \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{\boldsymbol{X}}}}$ *and* $FindSep(G, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{I}, \boldsymbol{E}) \neq \perp$ **then**
7 $\quad\quad$ **if** $\boldsymbol{I} = \boldsymbol{E}$ **then**
8 $\quad\quad\quad$ Output$(\mathbf{I} \setminus \mathbf{R})$
9 $\quad\quad$ **else**
10 $\quad\quad\quad$ W $\leftarrow$ arbitrary variable from $\mathbf{E} \setminus (\mathbf{I} \cup \mathbf{R})$
11 $\quad\quad\quad$ **if** $W \in \boldsymbol{V}_o$ **then**
12 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I} \cup \{W\}, \mathbf{E})$
13 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I}, \mathbf{E} \setminus \{W\})$
14 $\quad\quad\quad$ **if** $W \in \boldsymbol{V}_m$ *and* $\boldsymbol{R}_W \in \boldsymbol{E}$ **then**
15 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I} \cup \{W, \mathbf{R}_W\}, \mathbf{E})$
16 $\quad\quad\quad\quad$ ListSepConditions$(G, \mathbf{X}, \mathbf{Y}, \mathbf{R}, \mathbf{V}_o, \mathbf{V}_m, \mathbf{I}, \mathbf{E} \setminus \{W, \mathbf{R}_W\})$

$\boldsymbol{Z} \in \{\boldsymbol{Z} \mid (\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{R_Z}, \boldsymbol{R}_{\boldsymbol{X}\cap\boldsymbol{V}_m}, \boldsymbol{R}_{\boldsymbol{Y}\cap\boldsymbol{V}_m})_{G^{pbd}_{\boldsymbol{X},\boldsymbol{Y}}}$ & $(\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{R_Z} \mid \boldsymbol{X})_{G_{\overline{X}}}$ & $((\boldsymbol{X} \cap An(\boldsymbol{R_Z})) \perp\!\!\!\perp \boldsymbol{Y})_{G_{\underline{X}}}$ & $\boldsymbol{I} \subseteq \boldsymbol{Z} \subseteq \boldsymbol{E}\}$

Where $\boldsymbol{R_Z}$ *is a shorthand for* $\boldsymbol{R}_{\boldsymbol{Z}\cap\boldsymbol{V}_m}$.

ListSepConditions, by considering both including and not including each variable, recursively generates all subset of variables in $\mathbf{V}$ and for each generated set examines whether the conditions (b), (c), and (d) in Def. 4.1 holds or not. If those conditions were satisfied, the algorithm will return that candidate set as a m-adjustment set. ListSepConditions generates each potential set by taking advantage of back-track algorithm and at each recursion for a variable $W \in \mathbf{V}$ examines two cases of having W in candidate set or not. If $W \in \mathbf{V}_o$, then the algorithm examines having and not having this variable in the m-adjustment set and continues to decide about the rest of the variables in next recursion. If $W \in \mathbf{V}_m$, then the algorithm includes both W and R_W in the candidate m-adjustment set. Therefore, the algorithm considers both cases of having W, R_W and not having them in the candidate set. ListSepConditions, at the beginning of each recursion in Line 7, examines whether the candidate m-adjustment set so far satisfies the conditions (b), (c), (d) in Def. 6 or not. If any of them is not satisfied, the recursion stops for that candidate set. The function FindSep examines the existence of a set containing all variables in $\mathbf{I}$ and not having any of $\mathbf{V} \setminus \mathbf{E}$ that d-separates $\mathbf{X}$ from $\mathbf{Y}$. If this set does not exist FindSep returns $\perp$. ListSepConditions utilizes FindSep in order to check satisfaction of condition (b) in Def. 4.1 for the candidate set. Since the graph G that is given to FindSep is a proper back-door graph, all paths between $\mathbf{X}$ and $\mathbf{Y}$ in this graph is non-causal. Therefore, if a set separates $\mathbf{X}$ and $\mathbf{Y}$ in G^{pbd}, this set blocks all non-causal paths from $\mathbf{X}$ to $\mathbf{Y}$ in G.

The following theorem states that ListMAdj lists all the m-adjustment sets in a given m-graph G for recovering the causal effect of $\mathbf{X}$ on $\mathbf{Y}$.

Theorem 3 (Correctness of ListMAdj) *Given a m-graph G and sets of disjoint variables $\boldsymbol{X}$, $\boldsymbol{Y}$, ListMAdj returns all the sets that satisfy the m-adjustment criterion relative to $(\boldsymbol{X}, \boldsymbol{Y})$.*

The follow results state that Algorithm 1 is polynomial delay.

Proposition 4 (Complexity of ListSepConditions) *ListSepConditions for a given graph G has a time complexity of $O(n(n+m))$ polynomial delay where n and m are the number of variables and edges in G respectively.*

Theorem 4 (Complexity of ListMAdj) *ListMAdj for a given graph G returns all the m-adjustment sets with $O(n(n+m))$ polynomial delay where n and m are the number of variables and edges in G respectively.*

4.2 Finding minimum M-adjustment set

The problem of finding a m-adjustment set with minimum number of variables is important from several aspects. This can reduce the computational time, while making the result more interpretive. The cost of collecting more variables might be another reason researchers prefer to find a minimum set. Next we present an algorithm that for a given graph G with disjoint sets $\mathbf{X}$ and $\mathbf{Y}$ returns a m-adjustment set with the minimum number of variables.

Function FindMinAdjSet takes a m-graph G as input and returns a m-adjustment set with minimum number of variables. The function works by first removing all variables that violate Conditions (a), (c), and (d) in the m-adjustment criterion Def. 6 in lines 2 to 5, and then calling an external function FinMinCostSep given in van der Zander et al. (2014) which returns a minimum weight separator. FindMinAdjSet sets all the weights for each variable to be 1 to get a set with minimum size.

Algorithm 2: Find minimum size m-adjustment set

1 **Function** *FindMinAdjSet(*G, $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{V}_o$, $\boldsymbol{V}_m$, $\boldsymbol{R}$*)*

2 G′← compute proper back-door graph $G^{pbd}_{X,Y}$

3 $\mathbf{E} \leftarrow (\mathbf{V}_o \cup \mathbf{V}_m) \setminus \{\mathbf{X} \cup \mathbf{Y} \cup D_{pcp}(\mathbf{X}, \mathbf{Y})\}$.

4 $\mathbf{E}' \leftarrow \{E \in \mathbf{E} \mid E \in \mathbf{V}_o \text{ or } E \in \mathbf{V}_m \text{ and } (\mathbf{R}_E \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X})_{G'_{\overline{X}}}\}$

5 $\mathbf{E}'' \leftarrow \{E \in \mathbf{E}' \mid \mathbf{E} \in \mathbf{V}_o \text{ or } E \in \mathbf{V}_m \text{ and } (\mathbf{X} \cap An(\mathbf{R}_{\mathbf{E}}) \perp\!\!\!\perp \mathbf{Y})_{G'_{\underline{X}}}\}$

6 $\mathbf{W} \leftarrow 1$ for all variables

7 $\mathbf{I} \leftarrow$ empty set

8 $\mathbf{N} \leftarrow$ FindMinCostSep(G', $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{I}$, $\mathbf{E}''$, $\mathbf{W}$)

9 return $\mathbf{N} \cup \mathbf{R}_N$

Theorem 5 (Correctness of FindMinAdjSet) *Given a m-graph G and disjoint sets of variables $\boldsymbol{X}$, $\boldsymbol{Y}$, FindMinAdjSet returns a m-adjustment set relative to $(\boldsymbol{X}, \boldsymbol{Y})$ with minimum number of variables.*

Theorem 6 (Time Complexity of FindMinAdjSet) *FindMinAdjSet has a time complexity of $O(n^3)$.*

CHAPTER 5. SUMMARY AND DISCUSSION

In this thesis we introduce a m-adjustment formula for recovering causal effect in the presence of MNAR data and provide a necessary and sufficient graphical condition - m-adjustment criterion for when a set of covariates are valid m-adjustment. We introduce a ms-adjustment formulation for causal effects identification in the presence of both selection bias and MNAR data and provide a necessary and sufficient graphical condition - ms-adjustment criterion for when a set of covariates are valid ms-adjustment. We develop an algorithm that lists all valid m-adjustment or ms-adjustment sets in polynomial delay time, and an algorithm that finds a valid m-adjustment or ms-adjustment set containing the minimum number of variables. The algorithms are useful for data scientists to select adjustment sets with desired properties (e.g. low measurement cost). Adjustment is the most used tool for estimating causal effect in the data sciences. The results in this thesis should help to alleviate the problem of missing data and selection bias in a broad range of data-intensive applications. It is important to mention that in this paper we proposed an approach to recover all the causal effects queries that are recoverable by m-adjustment. However, it is possible that a causal effect is not obtainable by m-adjustment but can be obtain from other approaches. The problem of recovering causal effect in missing data by itself is still open. Future works can be done in finding complete algorithm for recovering the causal effects on missing data in any m-graphs with a more relaxed condition.

REFERENCES

Allison, P. D. (2001). *Missing data*, volume 136. Sage publications.

Bareinboim, E. and Tian, J. (2015). Recovering causal effects from selection bias. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, pages 3475–3481.

Bareinboim, E., Tian, J., and Pearl, J. (2014). Recovering from selection bias in causal and statistical inference. In *Proceeding of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 2410–2416.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Correa, J. D. and Bareinboim, E. (2017). Causal effect identification by adjustment under confounding and selection biases. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 3740–3746.

Correa, J. D., Tian, J., and Bareinboim, E. (2018). Generalized adjustment under confounding and selection biases. In *Thirty-Second AAAI Conference on Artificial Intelligence*, pages 6335–6342.

Huang, Y. and Valtorta, M. (2006). Identifiability in causal bayesian networks: A sound and complete algorithm. In *Proceedings of the 21st National Conference on Artificial Intelligence*, volume 2, pages 1149–1154. AAAI Press.

Koller, D., Friedman, N., and Bach, F. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Little, R. J. A. and Rubin, D. B. (1986). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc.

Mohan, K. and Pearl, J. (2014). Graphical models for recovering probabilistic and causal queries from missing data. In *Advances in Neural Information Processing Systems*, pages 1520–1528.

Mohan, K., Pearl, J., and Tian, J. (2013). Graphical models for inference with missing data. In *Advances in neural information processing systems*, pages 1277–1285.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition.

Perkovic, E., Textor, J., Kalisch, M., and Maathuis, M. H. (2017). Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *The Journal of Machine Learning Research*, 18(1):8132–8193.

Rubin, D. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Rubin, D. B. (1978). Multiple imputations in sample surveys-a phenomenological bayesian approach to nonresponse. In *Proceedings of the survey research methods section of the American Statistical Association*, volume 1, pages 20–34.

Shpitser, I., Mohan, K., and Pearl, J. (2015). Missing data as a causal and probabilistic problem. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 802–811.

Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 1219.

Shpitser, I., VanderWeele, T., and Robins, J. M. (2010). On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 527–536. AUAI Press.

Stekhoven, D. J. and Bühlmann, P. (2011). Missforest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118.

Takata, K. (2010). Space-optimal, backtracking algorithms to list the minimal vertex separators of a graph. *Discrete Applied Mathematics*, 158(15):1660–1667.

Tian, J. (2017). Recovering probability distributions from missing data. In *Proceedings of the Ninth Asian Conference on Machine Learning*, volume PMLR 77.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In *Eighteenth National Conference on Artificial Intelligence*, pages 567–573.

van der Zander, B., Liśkiewicz, M., and Textor, J. (2014). Constructing separators and adjustment sets in ancestral graphs. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 907–916. AUAI Press.

APPENDIX A. PROOFS IN CHAPTER 4

Proposition 2. Definition 4.1 and Definition 6 are equivalent.

Proof: Condition (c) and (d) in both definitions are the same. Condition (a) in Def. 4.1 indicates that $\mathbf{Z}$ cannot be in Dpcp($\mathbf{X}$,$\mathbf{Y}$). i.e., $\mathbf{Z}$ may not be descendant of any variables lies in proper causal path from $\mathbf{X}$ to $\mathbf{Y}$. This is as same as condition (a) in Def. 6. In order to prove Def. 4.1 $\rightarrow$ Def. 6, it is left to show Def. 4.1 leads to condition (b) in Def.6. By contradiction, assume there is a open non-causal path from a $X' \in \mathbf{X}$ to $Y' \in \mathbf{Y}$. Condition (b) in Def. 4.1 requires all non-causal proper back-door paths to be blocked. Therefore, This open non-causal path p does not belongs to proper back-door graph. The path p has edges coming out of $\mathbf{X}$ and belongs to a proper path q. Without lose of generality, assume $X' \in \mathbf{X}$ is the first and only variable in $\mathbf{X}$ that lies in the path p, otherwise, consider part of the path p with only X' at the beginning of it. Let W be the variable on the other side of this edge, and $Y' \in \mathbf{Y}$ be the last variable in path p and Y'' be the last one in q. Path p cannot be a direct path from X' to Y' since it is a non-causal path. Therefore, p should have colliders belong to $\mathbf{Z} \cup \mathbf{R_W}$. These colliders cannot belong to $\mathbf{Z}$ due to condition (a) in Def. 4.1. Consequently, they should belong to $\mathbf{R_W}$ which violates condition (c). Therefore, our assumption about the existence of the path p is not true. For the other direction, the closeness of all non-causal paths by $\mathbf{Z} \cup \mathbf{R_W}$ leads to closeness of non-causal proper back-door paths. To prove this theorem, we used some achievements in Correa et al. (2018).

Proposition 3 (Correctness of ListSepCondition). Given a m-graph G and sets of disjoint variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{E}$ and $\mathbf{I} \subseteq \mathbf{E}$, ListSepConditions lists all $\mathbf{Z}$ variables such that:

$$\mathbf{Z} \in \{\mathbf{Z} \mid (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}, \mathbf{R_Z}, \mathbf{R}_{\mathbf{X}\cap\mathbf{V}_m}, \mathbf{R}_{\mathbf{Y}\cap\mathbf{V}_m})_{G^{pbd}_{\mathbf{X},\mathbf{Y}}} \;\&\; (\mathbf{Y} \perp\!\!\!\perp \mathbf{R_Z} \mid \mathbf{X})_{G_{\overline{X}}} \;\&\; ((\mathbf{X} \cap An(\mathbf{R_Z})) \perp\!\!\!\perp \mathbf{Y})_{G_{\underline{X}}} \;\&\; \mathbf{I} \subseteq \mathbf{Z} \subseteq \mathbf{E}\}$$

Where $\mathbf{R_Z}$ is a shorthand for $\mathbf{R}_{\mathbf{Z}\cap\mathbf{V}_m}$.

Proof: The proof for this theorem includes two parts. In the first part, we prove the algorithm returns sound results, and in the second part we prove the algorithm returns all the correct results.

Part 1: Line 8 is where the algorithm returns the output. To get to line 8, the conditions in line 7 need to be satisfied. The conditions of $(\mathbf{Y} \perp\!\!\!\perp \mathbf{R_Z} \mid \mathbf{X})_{G_{\overline{X}}}$ and $((\mathbf{X} \cap An(\mathbf{R_Z})) \perp\!\!\!\perp \mathbf{Y})_{G_{\underline{X}}}$ are exactly checked in line 7. We explain how the algorithm makes sure the condition $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}, \mathbf{R_Z}, \mathbf{R}_{\mathbf{X}\cap\mathbf{V}_m}, \mathbf{R}_{\mathbf{Y}\cap\mathbf{V}_m})_{G^{pbd}_{\mathbf{X},\mathbf{Y}}}$ holds. Function FindSep examines if a candidate set is a valid separator for the sets $\mathbf{X}$ and $\mathbf{Y}$ in the graph G. Note that in our case, we are giving proper back-door graph as an input to this function. Therefore, all paths from $\mathbf{X}$ to $\mathbf{Y}$ are non-causal paths, and a set is a separator relative to the graph G and sets $\mathbf{X}$ and $\mathbf{Y}$, if and only if it closes all non-casual paths. If the set closes all non-causal path, the FindSep function returns true. Therefore, all outputs satisfy the three m-adjustment criterion (b,c,d).

Part 2: We prove the algorithm returns all sets satisfying m-adjustment criterion (b,c,d). The algorithm examines all subsets of $\mathbf{E}$ as a candidate sets by checking the two potential sets including and excluding $W \in \mathbf{E}$ in the sets with a backtracking. After selecting W, the algorithm evaluates type of W to see whether it belongs to $\mathbf{V}_m$ or $\mathbf{V}_o$. If $W \in \mathbf{V}_o$, the algorithm goes to the two next recursions of having W in the set and not having it. If $W \in \mathbf{V}_m$, it ensures to include R_W or not include it along variable W. Therefore, we evaluate all subsets of $\mathbf{E}$. It is only necessary for the algorithm to ensure not abort any recursion that is creating a valid m-adjustment sets. The only part of the algorithm that is responsible for aborting the recursion is line 7. ListSepCondition starts with a small set in each recursion path and in each run adds a variable to the set $\mathbf{I}$, if any of independencies $(\mathbf{Y} \perp\!\!\!\perp \mathbf{R_I} \mid \mathbf{X})_{G_{\overline{X}}}$ and $((\mathbf{X} \cap An(\mathbf{R_I})) \perp\!\!\!\perp \mathbf{Y})_{G_{\underline{X}}}$ in line 7 do not hold at any step of recursion, it means by adding more variables to $\mathbf{I}$ the dependency status wont change. Also, if FindSep cannot find an adjustment set for a given $\mathbf{I}$ and $\mathbf{E}$, then there is not any set having $\mathbf{I}$ as a subset of it that is adjustable. If a m-adjustment set blocks all non-causal paths, which means being separator, the FindSep should not return null for it. Therefore, the algorithm returns all sets **Z** satisfying the conditions and is correct.

Theorem 3(Correctness of ListMAdj). Given a m-graph G and sets of disjoint variables **X**, **Y**, ListMAdj returns all the sets that satisfy the m-adjustment criterion relative to (**X**,**Y**).

Proof: ListMAdj function in the first line excludes all variables violating condition (a) in m-adjustment criterion in Def. 6 and then calls ListSepCondition. Based on the theorem 3, It is proved that ListSepCondition returns all candidates for m-adjustment sets satisfying m-adjustment criterion (b,c,d) in Def. 6. Therefore, the function returns all sets that satisfying m-adjustment criterion.

Proposition 4 (Complexity of ListSepConditions). ListSepConditions for a given graph G has a time complexity of $O(n(n+m))$ polynomial delay where n and m are the number of variables and edges in G respectively.

Proof: To show the algorithm has a polynomial delay time complexity, we first demonstrate that it has an exponential time complexity and then we show it returns the first output as well as any two consecutive output in polynomial delay time. This algorithm examines all subset of variables in **V** as candidate m-adjustment sets. The number of subsets is exponential to the size of **V**. Therefore, algorithm has an exponential time complexity. Consider the recursion tree of ListSepConditions function. For each node in this tree, the function checks the two independencies mentioned in line 7 and then calls FindSep function. If all conditions in line 7 satisfy, the algorithm goes to the next node in recursion. Checking the first two conditions requires $O(n+m)$ and FindSep has a time complexity of $O(n+m)$. Therefore the time needed for examining each node is $O(3(n+m)) = O((n+m))$. In order to print an output the recursion needs to reach to the leaf of the tree. Recursion at each step removes a variable from the potential variables that are in the set. The depth of the tree is equal to size of $n = \mid \mathbf{V} \mid$. Therefore, the time needed to reach to the end of the recursion and return the output is $O(n(n+m))$. For generating the next output, the algorithm needs to goes back from a leaf node to the next leaf node. In the worst case, consider all branched were aborted due to rejection of any of conditions in line 7. In this case, the algorithm needs to check all the nodes from the end of the tree to top of it. This is equal to depth of the tree. A tree can have a depth with a length of at most n nodes. Therefore, generating the next output takes at most $O(n(n+m))$ time.

Theorem 4 (Complexity of ListMAdj). ListMAdj for a given graph G returns all the m-adjustment sets with $O(n(n+m))$ polynomial delay where n and m are the number of variables and edges in G respectively.

Proof: This function in the first part computes $D_{pcp}(\mathbf{X}, \mathbf{Y})$. This can be done in polynomial time. Later, the function calls ListSepCondition function which has exponential time complexity with $O(n(n+m))$ polynomial delay. Therefore, the time complexity of entire algorithm is $O(n(n+m))$ polynomial delay.

Theorem 5 (Correctness of FindMinAdjSet). Given a m-graph G and disjoint sets of variables $\mathbf{X}$, $\mathbf{Y}$, FindMinAdjSet returns a m-adjustment set relative to ($\mathbf{X}$,$\mathbf{Y}$) with minimum number of variables.

Proof: To prove FindMinAdjSet works properly, we prove this algorithm returns a valid m-adjustment and this m-adjustment has a minimum size. The algorithm excludes all variables that violate condition (a,c,d) in m-adjustment criterion in Def. 6. This function then find a minimum set $\mathbf{D}$ in a proper back-door graph $G^{pbd}_{X,Y}$ by using a FindMinCostSep. Since a separator in a $G^{pbd}_{X,Y}$, blocks all non-causal path, the returned set will satisfies the m-adjustment condition (b). Note that it might be thought that adding $\mathbf{R_D}$ variables to $\mathbf{D}$ will open a blocked path. However, no $\mathbf{R_D}$ lies on causal and non-causal path to $\mathbf{Y}$ because of independecy condition between $\mathbf{Y}$ and $\mathbf{R_D}$. Therefore, this situation does not happen. Now we prove FindMinAdjSet returns the minimum size m-adjustment. It is proved that FinMinCostSep van der Zander et al. (2014) returns a minimum separator in a graph G. Finding the minimum weight m-adjustment set in m-graph and causal graph is similar. The only difference between them is that in m-graph we have $\mathbf{R}$ variables. We explained that no $\mathbf{R_D}$ lies on the path to $\mathbf{Y}$. Therefore, a set with $\mathbf{R}$ variables as m-adjustment set cannot have the minimum size and it won't be returned by FindMinCostSep function.

Theorem 6 (Time Complexity FindMinAjdSet). FindMinAdjSet has a time complexity of O(n^3).

Proof: Consider a given graph $G = (\mathbf{V}, \mathbf{E})$ with $|V| = n$ and $|E| = m$. Generating proper back-door graph from G can be done in $O(m + n)$. The time complexity of computing D_{pcp} is $O(n(n + m))$ also. Therefore, Line 3 has a complexity of $O(n(n + m))$. Testing the d-separation can be done in $O(n + m)$ and since we are checking d-separation for all nodes the line 4 and 5 has $On(n + m)$ complexity. The time complexity of FindMinCostSep is $O(n^3)$. Therefore, the time complexity of FindMinAjdSet is $O(n^3)$.

APPENDIX B. PROOF FOR THEOREM 1

In the following section we used some of the proofs in (Correa and Bareinboim, 2017).

Lemma 2 *Let $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}$ be three disjoint sets of variables in an m-graph G. If a set $\boldsymbol{Z}$ satisfies the conditions in Def. 6 for a given set of treatment and outcome $\{\boldsymbol{X},\ \boldsymbol{Y}\}$, $\boldsymbol{Z}$ can be partitioned into the sets bellow:*

- $\mathbf{Z}_{nd}^{Y,1} = \{Z \mid Z \in \mathbf{Z} \setminus De_{\mathbf{X}} \text{ and } (Z \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{R_W})_{G_{\overline{X}}}\}$
- $\mathbf{Z}_{nd}^{X,1} = \{Z \mid Z \in \mathbf{Z} \setminus De_{\mathbf{X}} \setminus \mathbf{Z}_{nd}^{Y,1} \text{ and } (Z \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}_{nd}^{Y,1}, \mathbf{R_W})_{G_{\overline{\mathbf{X}(\mathbf{R_W})}}}\}$
- $\mathbf{Z}_{d}^{Y} = \{Z \mid Z \in \mathbf{Z} \cap\ De_{\mathbf{X}} \text{ and } (Z \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W})_{G_{\overline{X}}}\}$
- $\mathbf{Z}_{d}^{X} = \{\mathbf{Z} \cap De_{\mathbf{X}} \setminus \mathbf{Z}_{d}^{Y}\}$
- $\mathbf{Z}_{nd}^{Y,2} = \{Z \mid Z \in \mathbf{Z} \setminus\ De_{\mathbf{x}} \setminus \mathbf{Z}_{nd}^{Y,1} \setminus \mathbf{Z}_{nd}^{X,1} \text{ and } (Z \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{R_W})_{G_{\overline{X}}}\}$
- $\mathbf{Z}_{nd}^{X,2} = \mathbf{Z} \setminus De_{\mathbf{X}} \setminus \mathbf{Z}_{nd}^{Y,1} \setminus \mathbf{Z}_{nd}^{X,1} \setminus \mathbf{Z}_{nd}^{Y,2}$

Based on this partitioning, the following independencies can be conclude: $(\mathbf{Z}_{d}^{X} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W})_{G_{\overline{X(\mathbf{Z}_{d}^{Y}, \mathbf{R_W})}}}$ and $(\mathbf{Z}_{nd}^{X,2} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \setminus \mathbf{Z}_{nd}^{X,2}, \mathbf{R_W})_{G_{\overline{X(\mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{R_W})}}}$.

Proof, Part 1. To prove this independency holds: $(\mathbf{Z}_{nd}^{X,2} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \setminus \mathbf{Z}_{nd}^{X,2}, \mathbf{R_W})_{G_{\overline{X(\mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{R_W})}}}$, we assume, by contradiction, that this assumption is not true. Therefore, there should be an open path between $Z' \in \mathbf{Z} \setminus\ De_{\mathbf{X}} \setminus \mathbf{Z}_{nd}^{Y,1} \setminus \mathbf{Z}_{nd}^{X,1} \setminus \mathbf{Z}_{nd}^{Y,2}$ and $X \in \mathbf{X}$. We name this path q in the graph $G_{\overline{X}}$. Since Z' does not belong to $\mathbf{Z}_{nd}^{Y,1}$ and based on the definition of $\mathbf{Z}_{nd}^{Y,1}$, there exists an open path between $Y \in \mathbf{Y}$ and Z'. We call this path p. The only collider that is allowed to exist in path p is Z'. P cannot have any variable as a collider in $\{\mathbf{R_W}\}$ due to condition (c) that requires the d-separation between Y and $\{\mathbf{R_W}\}$ for a given X. The variable Z' is not in $\mathbf{Z}_{nd}^{Y,2}$ based on the definition of $\mathbf{Z}_{nd}^{X,2}$.

Therefore, p does not contain any covariate in $\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}$ or $\mathbf{Z}_{nd}^{Y,2}$; Otherwise, these sets close p and lead Z' belongs to $\mathbf{Z}_{nd}^{Y,2}$ as the fact that p does not have any colliders. We have two situations: X is or is not the ancestor of $\{\mathbf{R_W}\}$. In the first scenario, the arrow in path q needs to come out of X. The definition of Z' necessitates that Z' is not a descendant of X. Therefore, there will be colliders in q. Due to the assumption that q is open, these colliders must be ancestors of $\{\mathbf{R_W}\}$. This is in contradiction with the assumption that X is not an ancestor of the variables in $\{\mathbf{R_W}\}$. For the case that arrows coming into X, consider the joint path p and q. In this path, X should be an ancestor of $\{\mathbf{R_W}\}$ which is in contradiction to condition (d). If Z' is a collider in the joint path, we will have a non-causal open path which is against condition (b). If the arrows come out of X in path q, due to the fact that Z' is non-descendant of X, we need to have a collider in q. Based on our assumption q is open. Therefore, the collider belongs to $\{\mathbf{R_W}\}$, otherwise Z' would be in the set $\mathbf{Z}_{nd}^{X,1}$. Having $W \in \{\mathbf{R_W}\}$ as a collider in q necessitate Z' to be a collider by itself since we need to close the open path from W to Y based on condition (c). However, then conditioning on Z' will open the non-causal path from $\mathbf{X}$ to $\mathbf{Y}$. Therefore, the assumption of the existence of such Z' is invalid.

Proof, Part 2. To prove this independency, $(\mathbf{Z}_{d}^{X} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W})_{G_{\overline{\mathbf{X}(\mathbf{Z}_{d}^{Y},\mathbf{R_W})}}}$, we consider two cases of $\mathbf{Z}_{d}^{X} \neq 0$ and $\mathbf{Z}_{d}^{X} = 0$. For the first case, by contradiction, assume the independency is not true. Therefore, there exists an open path q between X and $Z' \in \mathbf{Z}_{d}^{X}$, while the rest of $\mathbf{Z}_{d}^{Y}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W}$ are observed. Since Z' belongs to $\mathbf{Z}_{d}^{X}$, all variables in path q must be descendant of X. We know that $Z' \notin \mathbf{Z}_{d}^{Y}$ based on the definition $\mathbf{Z}_{d}^{X}$. Therefore, there exists an open path p from Z' to Y while the variables $\mathbf{X}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{R_W}$ are observed. There is no variable belongs to $\mathbf{X}$ in path p based on the condition (b).

Consider the junction of paths q and p. Path p cannot be directed since this junction path will be a proper causal path with some nodes from Z' on it. This is against condition (a). Therefore, p needs to have colliders on it. Based on condition (b), Z' cannot be collider, unless the path q be closed by $\mathbf{Z}_{nd}^{X,1}$ and $\mathbf{Z}_{nd}^{X,1}$ which, based on their definition, is not possible to have them on q. If

Z' is not collider in p, there needs to be another collider variable $W \in \{\mathbf{R_W}\} \cup \mathbf{Z}_{nd}^{Y,1} \cup \mathbf{Z}_{nd}^{X,1} \cup \mathbf{Z}_{d}^{Y}$. None of these three sets can be collider. $\{\mathbf{R_W}\}$ cannot be collider because of condition (c). $\mathbf{Z}_{d}^{Y}$ cannot be collider since it is independent of $\mathbf{Y}$. Lastly, the two sets $\mathbf{Z}_{nd}^{Y,1}$ or $\mathbf{Z}_{nd}^{X,1}$ are descendant of $\mathbf{X}$. Therefore, they cannot be used as W. Since there is no valid variable to be as collider in path p, our assumption of existence of path q is not a legitimate assumption.

Theorem 1 (M-Adjustment) A set $\mathbf{Z}$ is a m-adjustment set for recovering causal effect of $\mathbf{X}$ on $\mathbf{Y}$ by the m-adjustment formula in Def. 3.1 if and only if it satisfies the m-adjustment criterion in Def. 6.

Proof (if): Based on lemma 1, a valid adjustment set $\mathbf{Z}$ can be partitioned into the $\{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{Z}_{nd}^{Y,2}, \mathbf{Z}_{nd}^{X,2}\}$. Based on this fact, the casual effect of $\mathbf{X}$ on $\mathbf{Y}$ can be computed as follows:

According to condition (c), $(\mathbf{Y} \perp\!\!\!\perp \{\mathbf{R_W}\} \mid \mathbf{X})_{G_{\overline{X}}}$ and $\{\mathbf{R_W}\}$ can be inserted into the following expression:

$$P(\mathbf{y} \mid do(\mathbf{x})) = P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{R_W} = 1) \tag{B.1}$$

$\mathbf{Z}_{nd}^{Y,1}$ is independent of $\mathbf{Y}$. Therefore, it can be added to the first factor. Introducing the second factor with summation over $\mathbf{Z}_{nd}^{Y,1}$ values is valid.

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{Y,1} \mid \mathbf{R_W} = 1) \tag{B.2}$$

By conditioning on $\mathbf{Z}_{nd}^{X,1}$ in the first factor, we get the following expression:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{X,1} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{R_W} = 1)$$
$$P(\mathbf{Z}_{nd}^{Y,1} \mid \mathbf{R_W} = 1) \tag{B.3}$$

In the second factor we can remove do($\mathbf{x}$) based on the fact that $(\mathbf{Z}_{nd}^{X,1} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}_{nd}^{Y,1}, \mathbf{R_W})_{G_{\overline{X(\mathbf{R_W})}}}$ ($G_{\overline{\mathbf{X}(\mathbf{R_W})}} = G_{\overline{\mathbf{X}(\mathbf{Z}_{nd}^{Y,1}, \mathbf{R_W})}}$ since $\mathbf{Z}_{nd}^{Y,1}$ is independent of $\mathbf{X}$), we can use rule 3 of the do-calculus and remove do($\mathbf{x}$). Taking advantage of the chain rule, factors two and three can be joined.

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{Y,1} \mathbf{Z}_{nd}^{X,1} \mid \mathbf{R_W} = 1) \tag{B.4}$$

$\mathbf{Z}_d^Y$ is independent of $\mathbf{Y}$, so we can insert it in the first factor. $\mathbf{Z}_{nd}^{Y,1}$ can be inserted in the second factor and summed out on all its possible values.

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y \mid \mathbf{R_W} = 1) \tag{B.5}$$

Since $\mathbf{Z}_d^X$ is not independent of Y, conditioning on it leads to:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{Z}_d^X} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{Z}_d^X, \mathbf{R_W} = 1) P(\mathbf{Z}_d^X \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y \mid \mathbf{R_W} = 1) \tag{B.6}$$

do($\mathbf{x}$) in the second factor can be removed by using rule 3 of the do-calculus, since the following independency: $(\mathbf{Z}_d^X \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}_d^Y, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{R_W} = 1)_{G_{\overline{\mathbf{X}(\mathbf{Z}_d^Y, \mathbf{R_W})}}}$ holds. Then applying the chain rule on the factors two and three leads to the following expression.

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{Z}_d^X} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{Z}_d^X, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{Z}_d^X \mid \mathbf{R_W} = 1) \tag{B.7}$$

We use this independency:$(\mathbf{Z}_{nd}^{Y,2} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{X}, \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_d^Y, \mathbf{Z}_d^X, \mathbf{R_W} = 1)_{G_{\overline{X}}}$, and insert $\mathbf{Z}_{nd}^{Y,2}$ into the first factor. In the next step we add $\mathbf{Z}_{nd}^{Y,2}$ to the second factor and put a summation of it.

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{Z}_{nd}^{Y,2}} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{Z}_{d}^{Y,2}, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{Z}_{d}^{Y,2} \mid \mathbf{R_W} = 1) \tag{B.8}$$

We condition on $\mathbf{Z}_{nd}^{X,2}$ in the first factor:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{Z} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{Z}, \mathbf{R_W} = 1) P(\mathbf{Z}_{nd}^{X,2} \mid do(\mathbf{x}), \mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{Z}_{d}^{Y,2}, \mathbf{R_W} = 1)$$
$$P(\mathbf{Z}_{nd}^{Y,1}, \mathbf{Z}_{nd}^{X,1}, \mathbf{Z}_{d}^{Y}, \mathbf{Z}_{d}^{X}, \mathbf{Z}_{d}^{Y,2} \mid \mathbf{R_W} = 1) \tag{B.9}$$

By getting help from rule 3 of do-calculus and using this independency $(\mathbf{Z}_{nd}^{X,2} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z} \setminus \mathbf{Z}_{nd}^{X,2}, \mathbf{R_W} = 1)_{G_{\overline{\mathbf{X}(\mathbf{Z}_d^Y, \mathbf{Z}_d^X, \mathbf{R})}}}$ do($\mathbf{x}$) is removed from the second factor:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}} P(\mathbf{y} \mid do(\mathbf{x}), \mathbf{z}, \mathbf{R_W} = 1) P(\mathbf{z} \mid \mathbf{R_W} = 1) \tag{B.10}$$

Based on conditions (a , b) we have $(\mathbf{Y} \perp\!\!\!\perp \mathbf{X} \mid \mathbf{Z}, \mathbf{R_W} = 1)_{G_{\underline{X}}}$

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{Z}} P(\mathbf{y} \mid \mathbf{x}, \mathbf{z}, \mathbf{R_W} = 1) P(\mathbf{z} \mid \mathbf{R_W} = 1) \tag{B.11}$$

Proof (only if): In this part, we prove that if any of the criterion in m-adjustment criterion is not true, there will be a graph G that for a given set of treatment and outcome ($\mathbf{X}$,$\mathbf{Y}$), the causal effect of $P(\mathbf{y} \mid do(\mathbf{x}))$ is not recoverable. The condition (b) in m-adjustment criterion is the extended version of the condition (b) in adjustment set. The only difference is $\mathbf{R}$ is observed rather than only S. Therefore, we prove besides $\mathbf{Z}$, $\mathbf{R}$ are required to block all non-causal paths. By contradiction, assume this is not the case. Therefore, there should be a non-causal path q between X and Y, that is closed by observed $\mathbf{Z}$, and gets open when there is a condition on $\mathbf{R}$. In order to demonstrate that the graph G with this non-causal path is non-recoverable, we consider two models M_1 and M_2 both compatible with the graph G. We assign P_1 as a probability distubtion corresponding to M_1

and P_2 for M_2. M_1 and M_2 agree on probability distribution under selection, and MNAR biases and are disagree on the causal effect of the set of treatment on the set of outcome.

$$P_1(\mathbf{v} \mid \mathbf{R}^{\mathbf{v}} = 1) = P_2(\mathbf{v} \mid \mathbf{R}^{\mathbf{v}} = 1) \tag{B.12}$$

$$P_1(\mathbf{y} \mid do(\mathbf{x})) \neq P_2(\mathbf{y} \mid do(\mathbf{x})) \tag{B.13}$$

We construct M_1 in a way to be compatible with the graph $G_{\overline{\mathbf{R}_W}}$, separating all $\mathbf{R}_W$ from their parents, $(\mathbf{V} \perp\!\!\!\perp \mathbf{R}_W)_{M_1}$,and M_2 compatible with the graph G:

$$P_1(\mathbf{v} \mid \mathbf{R}^{\mathbf{v}} = 1) = P_1(\mathbf{v} \mid \mathbf{R}^{\mathbf{v}} \setminus \{\mathbf{R}_{\mathbf{W}} = 1\}) \tag{B.14}$$

The causal effect query needs to be recoverable for any parametrization of probability distributions P_1, P_2. We construct P_2 in a way that equation B.12 holds.

Without loss of generality, we are considering the path between $Y' \in \mathbf{Y}$ and $X' \in \mathbf{X}$ that condition (b) of m-adjustment criterion does not satisfy in it. Therefore, our desired model will have all the variables in the rest of the graph d-separated from the variables in the path. We have:

$$P(\mathbf{y} \mid do(\mathbf{x}')) = \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}', \mathbf{z}, \mathbf{R}_W = 1) P(\mathbf{z} \mid \mathbf{R}_W = 1) \tag{B.15}$$

$$= \prod_{\mathbf{Y}} \sum_{\mathbf{z}} P(\mathbf{y} \mid \mathbf{x}', \mathbf{z}, \mathbf{R}_W = 1) P(\mathbf{z} \mid \mathbf{R}_W = 1) \tag{B.16}$$

$$= (\prod_{\mathbf{Y} \setminus Y'} P(\mathbf{y})) \sum_{\mathbf{z}} P(\mathbf{y}' \mid \mathbf{x}', \mathbf{z}, \mathbf{R}_W = 1) P(\mathbf{z} \mid \mathbf{R}_W = 1) \tag{B.17}$$

$$= \gamma \sum_{\mathbf{z}} P(\mathbf{y}' | \mathbf{x}', \mathbf{z}, \mathbf{R}_W = 1) P(\mathbf{z} \mid \mathbf{R}_W = 1) \tag{B.18}$$

γ in above expression indicates product of marginal distribution $\mathbf{Y} \setminus Y'$.

The open non-causal path between X' and Y' that is blocked by $\mathbf{Z}$ but opened with $\mathbf{R}' \subseteq \mathbf{R}_{\mathbf{W}} = 1$, needs $\mathbf{R}'$ to be colliders. Fig. B.1 shows a general case for when the set $\mathbf{R}'$ has size 1. By a small change we will get Fig. B.2 which shows the general case for when $\mathbf{R}'$ has a size greater than one.

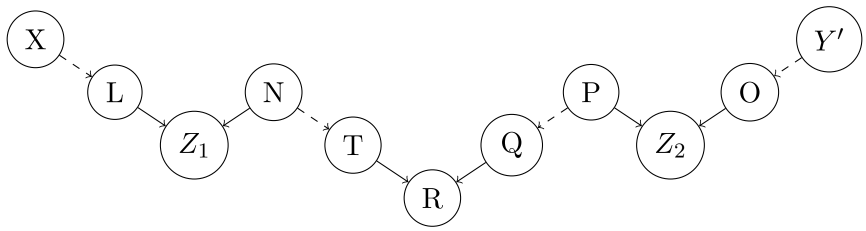


Figure B.1: This graph indicates an open non-causal path between X and Y with conditioning on $\mathbf{R}$ and $\mathbf{Z}$. The path from N to T and from P to Q can be substituted by a path with any number of $Z \in \mathbf{Z}$. Dotted edges refer to chains of nodes.

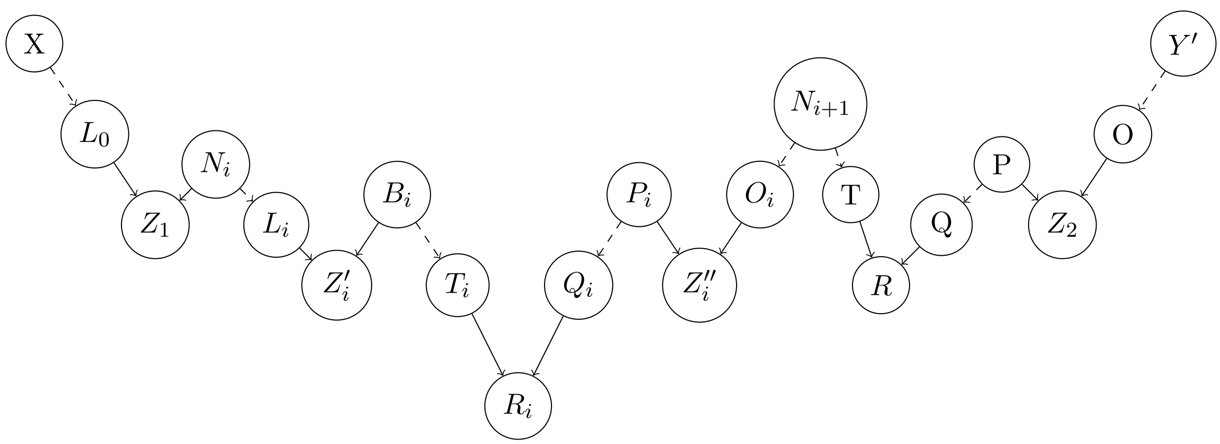


Figure B.2: The path from N_i to N_{i+1} can recursively be substituted by more of the path of the same kind to include arbitrary number of $\mathbf{R}$ variables. R_i and R indicate variables belong to $\mathbf{R_W}$.

case 1: There is only one collider belongs to $\mathbf{R}'$. The proof for this part is as same as the selection bias (Correa et al., 2018). Therefore, we omit repeating it.

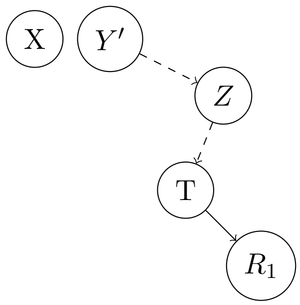
case 2: The set $\mathbf{R}'$ might have the size greater than one. Fig. B.2 expresses a graphical representation for this situation. To prove this case, we provide a parametrization for the path from N_i to N_{i+1}. The rest of the proof will be similar to case 1.

We assign $P_1(N_{i+1}) = P_1(P_i) = P_1(B_i) = P_1(N_i) = 1/2, P(O_i \mid N_{i+1}) = 1/2 + \epsilon_5/2, P(O_i \mid \overline{N_{i+1}}) = 1/2 - \epsilon_5/2, P(L_i \mid N_i) = 1/2 + \epsilon_6/2, P(L_i \mid \overline{N_i}) = 1/2 - \epsilon_6/2, P(T_i \mid B_i) = 1/2 + \epsilon_7/2, P(T_i \mid \overline{B_i}) = 1/2 - \epsilon_7/2, P(Q_i \mid P_i) = 1/2 + \epsilon_8/2, P(Q_i \mid \overline{P_i}) = 1/2 - \epsilon_8/2, P(z_i'' \mid P_i, O_i) = P(z_i'' \mid \overline{P_i}, O_i) = P(z_i'' \mid \overline{O_i}, P_i) = P(z_i'' \mid \overline{P_i}, \overline{O_i}) = 1/2, P(R_i \mid T_i, Q_i) = P(R_i \mid \overline{T_i}, Q_i) = P(R_i \mid \overline{T_i}, Q_i) = P(R_i \mid$

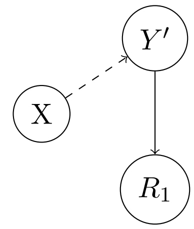
$\overline{T_i}, \overline{Q_i}) = 1/2, P(Z_i' \mid B_i, L_i) = P(Z_i' \mid \overline{B_i}, L_i) = P(Z_i' \mid \overline{B_i}, L_i) = P(Z_i \mid \overline{B_i}, \overline{L_i}) = 1/2$, Where $\epsilon_i = (\frac{1}{5}^{k_i})$, k_5 is the length of the path N_{i+1} to O_i, k_6 is the length of the path from N_i to L_i,k_7 is the length of the path from B_i to T_i, and k_8 is the length of the path from P_i to Q_i. This parametrization provides the same values for Q_1 and Q_2 as case 1.

Now we evaluate the necessity of condition (c). Fig. B.3 and Fig. B.4 show all the cases violating condition (c). Note that in these figures, $R_1, R_2, R_3, R_i \in \mathbf{R_W}$, and, by mentioning R_i, we are referring to have R_i violated condition (c). The proof for cases 1 to 6 is similar as (Correa et al., 2018). Cases of 7, 10, 11, and 12 are extended versions of case 2, and case 8, 13, and 14 are extended versions of case 3. It is clear that by adding more edges to case 5 we can obtain case 9. We can conclude these extended versions are not recoverable, since if recoverability is impossible in a graph, adding more edges does not change recoverability status.

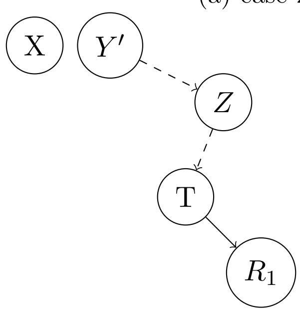
In this part we are evaluating whether the condition (d) is necessary or not. To condition (d) be violated, there should be a back-door path between $X \in \mathbf{X} \cap An(\mathbf{R_W})$ and $Y' \in \mathbf{Y}$. We name this path p. Based on condition (b), this path should be blocked by some $Z' \in \mathbf{Z}$. There are two scenarios. The first one is, we have a direct causal path from X to Y' and the second one is, lack of existence of that type of path, both cases are demonstrated in Fig. B.5. The proof for these cases are similar to selection bias one and is provided in (Correa et al., 2018).



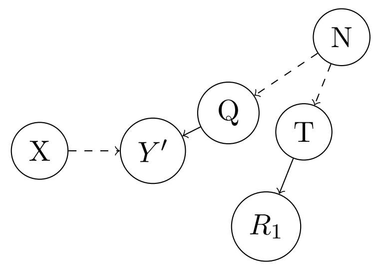
(a) case 2



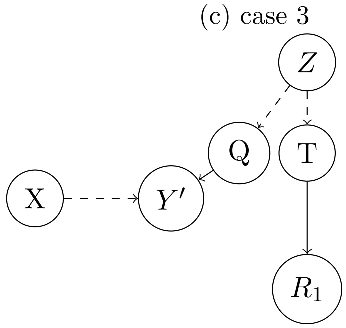
(b) case 1



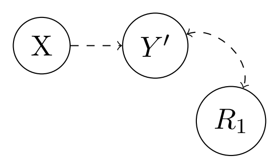
(c) case 3



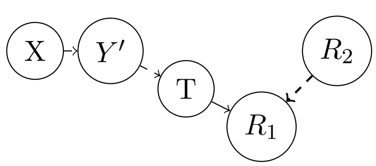
(d) case 4



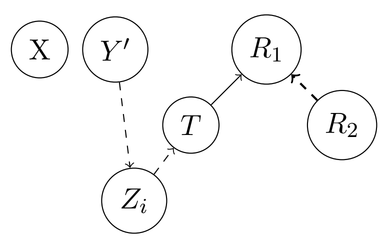
(e) case 5



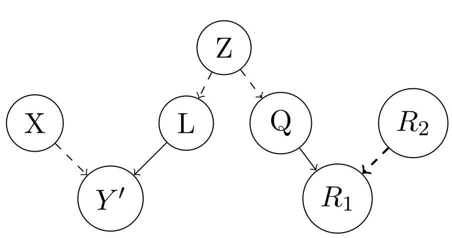
(f) case 6



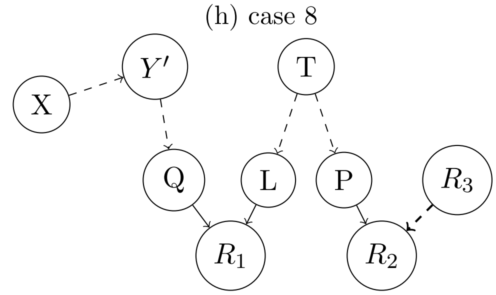
(g) case 7



(h) case 8

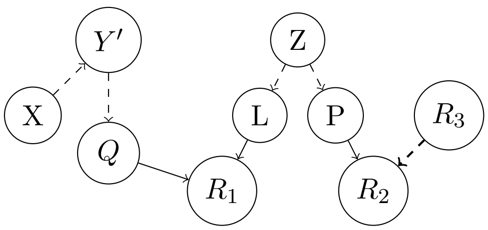


(i) case 9

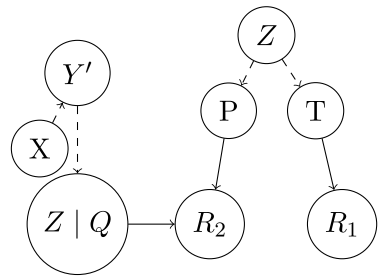


(j) case 10

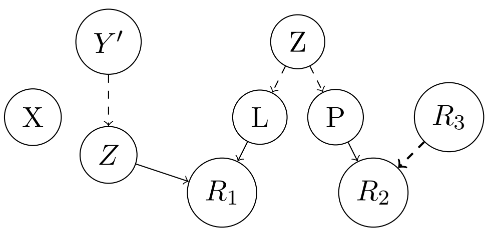
Figure B.3: Cases condition (c) in M-adjustment criteria are violated part 1.



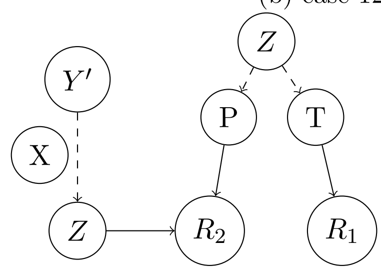
(a) case 11



(b) case 12

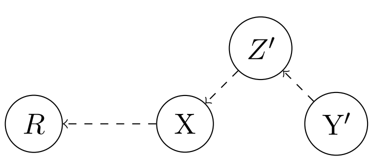


(c) case 13

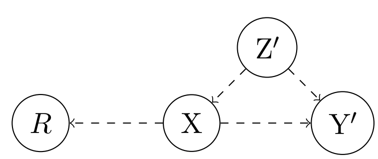


(d) case 14

Figure B.4: Cases condition (c) in M-adjustment criteria are violated part 2



(a) case 1



(b) case 2

Figure B.5: Cases considered for the necessity of condition (d) in M-adjustment criteria.